METHODS OF PRODUCING CELLS AND ANIMALS COMPRISING TARGETED GENE MODIFICATIONS AND COMPOSITIONS RELATING THERETO

RELATED APPLICATIONS

5   This application claims priority to U.S. Provisional Application No. 60/232,957, filed September 15, 2000.

FIELD OF THE INVENTION

The present invention is directed to compositions and methods relating to the generation of cells and animals comprising a genetic modification or alteration of a targeted gene.

10   ## BACKGROUND OF THE INVENTION

The ability to manipulate the mammalian genome, and in particular, the ability to develop animals with specific genes altered or inactivated has been invaluable to the study of gene function. The capability to modify or inactivate a gene can lead to unexpected discoveries of a gene and/or mechanisms responsible for disease with similar manifestations in humans. These

15   genetically engineered animals are also useful for testing drug treatments and developing gene therapy strategies. (*See, e.g.,* Bradley A., 1993, *Recent Prog. Horm. Res.* 48:237-251).

Mouse mutants have provided an extremely useful source of knowledge of mammalian development, cellular biology, and physiology, and have provided models for human diseases. An example of a well-known animal having a mutated or "knock-out" gene includes mice

20   carrying a specifically modified or disrupted form of a chloride-channel gene. These mice develop a disease closely resembling human cystic fibrosis. Other examples of mice that have proven to be particularly valuable include those with alterations of genes encoding lymphocyte-specific tyrosine kinase p56.sup.lck and Lyt-2, alpha.-Calcium Calmodulin kinase II gene, the C/EPB.alpha. gene, and the BAX gene. (*See, e.g.*, Snowouwaert *et al.*, 1992, *Science* 257:1083-

25   1088; Dorin *et al.*, 1992, *Nature* 359:211-215; U.S. Patent No: 5,625,122; U.S. Patent No. 5,530,178; Silva *et al.*, 1992, *Science*, 257:201; Wang *et al.*, 1995, *Science*, 269:1108; Knudsen *et al.*, 1995, *Science*, 270:960).

Determining how a gene functions ultimately requires genetic analysis *in vivo*. The mouse, for example, is a proven model system for studying various aspects of *in vivo* genetic

30   analysis and mammalian development. (*See, e.g.*, Paigen K., 1995, *Nature Med.* 1:215-220). Understanding how mammalian genes function, including genes from humans, has relied heavily

on gene targeting technologies. Gene targeting allows for the generation of mice with a specifically-altered genotype.

Genetically altering specifically-targeted DNA sequences within eukaryotic genomes relies on homologous recombination to replace normal gene sequences in a cell with modified exogenous sequences that introduce the desired mutation. Such targeted replacement of a DNA sequence occurs in only a small fraction of the treated cells, while the incoming DNA is subject most often to random integrations. (*See, e.g.,* Bollag *et al.,* 1989, *Annu. Rev. Genet.* 23:199-225). More particularly, exogenous sequences transferred into eukaryotic cells undergo homologous recombination with homologous endogenous sequences only at very low frequencies, and are so inefficiently recombined that large numbers of cells must be transfected, selected, and screened in order to generate a desired correctly targeted homologous recombinant. (*See, e.g.,* Kucherlapati *et al.,* 1984, *Proc. Natl. Acad. Sci. (U.S.A.)* 81: 3153; Smithies, O., 1985, *Nature* 317: 230; Song *et al.,* 1987, *Proc. Natl. Acad. Sci. (U.S.A.)* 84: 6820; Doetschman *et al.,* 1987, *Nature* 330: 576; Kim and Smithies, 1988, *Nucleic Acids Res.* 16: 8887; Shesely *et al.,* 1991, *Proc. Natl. Acad. Sci. (U.S.A.)* 88: 4294; Kim *et al.,* 1991, *Gene* 103: 227).

The most common approach to producing these transgenic animals involves the disruption of a target DNA sequence by insertion of a DNA construct encoding a selectable marker gene flanked by DNA sequences homologous to part of the target gene. When properly designed, the DNA construct effectively integrates into and disrupts the targeted gene via homologous recombination, thereby preventing the normal expression of an active gene product encoded by that gene.

Typically, gene targeting strategies employed to generate animals having specific mutations involve the following steps: 1) directed mutagenesis of the target gene *in vitro*; 2) introduction of the mutant gene into cultured embryonic stem cells; 3) screening for cell lines carrying the desired homologous recombination (*i.e.,* gene replacement) event; and 4) generation of mice that transmit the mutant gene. (*See, e.g.,* Capecchi, 1989, *Trends In Genetics* 5(3):70-76; Capecchi, 1989, *Science* 244(4910):1288-1292).

Directed mutagenesis of the target gene *in vitro* can be achieved using standard molecular biology and DNA cloning techniques. Typically, a functionally-relevant gene sequence is deleted and replaced with a selectable marker gene. The *neo* gene, which encodes neomycin phosphotransferase and confers cellular resistance to neomycin, G418 and related drugs, is

2

routinely used as the selectable marker gene. In general, the deletion and replacement of the functionally-relevant gene are designed to generate a null mutation in the target gene disrupting its normal activity or function.

To introduce a mutant gene into cultured embryonic stem cells, a genetic construct or targeting vector is grown as a DNA plasmid in bacteria and then transfected into murine embryonic stem cells *in vitro*. The desired transfected cells, which represent a small fraction of the total cell population, are purified from those that failed to take in the vector by positively selecting for the marker gene in the transfected cells. Specifically, addition of neomycin to the culture kills untransfected cells, thus, selecting for the outgrowth of resistant transfected cells that express the *neo* gene. These resistant cells grow into colonies, each representing clonal populations derived from independently transfected cells.

Screening for cell lines carrying the desired homologous recombination event allows for the identification of cells in which the specific gene replacement has occurred. Given that random integration typically occurs more frequently than does homologous recombination, only a small minority of the colonies will be derived from cells having homologous gene replacement. This screening process requires that DNA samples isolated from individual cell lines be analyzed for homologous recombination, usually by the polymerase chain reaction (PCR) or DNA blot hybridization (Southern blotting).

To generate mice that transmit the mutant gene, embryonic stem cells carrying the desired homologous recombination event can be injected into mouse blastocysts. The blastocysts are then implanted into pseudopregnant females to generate chimeric mice, comprised of both mutant and wild-type cells. If the germline has been populated with mutant cells, then the targeted allele can be transmitted to subsequent generations, and the phenotypic consequences of the mutation can be assessed.

One of the most challenging aspects in generating animals comprising a targeted gene modification is the identification and isolation of the rare cell line that carries the homologous recombination event. One approach to combating this difficulty involves the addition of a negative selection step. This technique allows for the enrichment of the transfected cell population for the desired cells, relying on negative selection to specifically kill cells that carry random integrations. (*See, e.g.*, U.S. Patent No.: 5,627,059). In addition to the general techniques described above, this positive/negative selection (PNS) method requires the cloning

3

of a negative selectable marker into the targeting vector and a further negative selection step. The gene encoding thymidine kinase (TK) is routinely used as the negative selection marker in the PNS method.

The PNS method involves a process in which a first drug is added to the cell population, for example, a neomycin-like drug to select for growth of transfected cells, *i.e.* positive selection. A second drug, such as FIAU is subsequently added to kill cells that express TK, *i.e.* negative selection. However, addition of the second drug can be quite toxic to the cells and may negatively affect the ability of the cells to populate the germline. (*See, e.g.,* Yanagawa et al., 1999, *Transgenic Research* 215-221). Unfortunately, in addition to homologous recombination, many random integration events will also inactivate TK. Indeed, although the negative selection enriches the cell population for homologous recombinants, this population still predominantly contains random integration events.

Mammalian cells have a remarkable ability to support nonhomologous recombination of incoming DNA. For example, animals bearing a foreign gene randomly inserted into their genome to express a foreign protein are reported in the art. These animals are most often used to produce, for example, a pharmaceutical substance. Typically, in this process expression of the foreign gene's coding sequence is under the control of a promoter.

Previous studies demonstrated that control of eukaryotic transcriptional promoters, can be modified to respond to bacterial transcription factors. (*See, e.g.,* Hu and Davidson, *Molecular and Cellular Biology* 10(12):6141-6151; Hu and Davidson, 1991, *Gene* 99(2):141-150; Hu and Davidson, 1987, *Cell* 48(4):555-566; Hu and Davidson, 1988, *Gene* 62(2):301-313; Hannan *et al.*, 1993, *Gene* 130(2):233-239).

However, the method of expressing a foreign gene of interest in a mammalian cell by randomly inserting the gene into the genome of the animal is contrary to the process of gene targeting. Gene targeting relies on homologous recombination, wherein the goal is to produce an animal carrying a modified or disrupted form of a specific gene of interest.

As described above, the experimental challenge in gene targeting lies in identifying the rare colonies of cells carrying the desired mutated target gene. As it is often difficult to differentiate between random insertions and homologous recombination, a need in the art exists for methods that enhance and promote the recovery of homologous recombination events, while

4

providing a faster, more efficient, and more reliable means for generating cells and animals having specific genes modified or disrupted.

SUMMARY OF THE INVENTION

The present invention relates to novel compositions and methods useful in the production of cells and animals having a genetic alteration or modification of a targeted DNA sequence. More particularly, the present invention provides compositions and methods that are capable of modifying a target gene in a cell with high efficiency and specificity.

The present invention provides a regulated positive selection vector (referred to herein as "targeting vector") that is capable of modifying or disrupting expression of a targeted gene. The targeting vector comprises a first sequence homologous to a portion or region of a target gene sequence and a second sequence homologous to a second portion or region of a target gene sequence. The targeting vector also includes a selectable marker cassette that comprises a selectable marker gene. Preferably, the selectable marker cassette is positioned in between the first and the second sequence homologous to a region or portion of the target gene sequence. In one aspect, the selectable marker cassette, in addition to a selectable marker gene, also comprises a sequence that initiates, directs, or mediates transcription of the selectable marker. The targeting vector also comprises a regulator that has the ability to control or regulate the expression of the selectable marker. Preferably, the regulator is positioned outside of the first or second sequence homologous to a region or portion of the target gene.

The present invention also provides novel methods of modifying a target gene. In one aspect, the present invention provides novel methods of producing cells having a disruption or modification of a target gene and generating animals comprising these genetic modifications. In accordance with this aspect, the targeting vector of the present invention is introduced into cells that are capable of homologous recombination. In this process, the transfected DNA will integrate or recombine with and replace the homologous portions of the endogenous sequence. When homologous recombination occurs between the homologous portions of the endogenous target gene, the targeting vector excluding the regulator is incorporated into the genome of the cell. However, most frequently the transfected DNA will integrate at a random site in the genome of the cell. In such a case, the targeting vector including the regulator is incorporated into a random site in the genome of the cell. The regulator inhibits or suppresses expression of the selectable marker, thus, if the regulator sequence is not incorporated into the genome of the

5

cell by way of homologous recombination, the selectable marker is expressed. Thus, cells wherein gene targeting has occurred can be selected by way of the selection marker only. As expression of the selectable marker is under the control of the regulator, cells wherein random integration occur, do not survive the addition of the selection agent, as the regulator incorporated into a random site in the genome of the cell, blocks or inhibits expression of the selectable marker gene.

In a further aspect, the present invention provides a method of identifying cells comprising the targeted gene modification. Furthermore, methods of the present invention provide a faster and more efficient means for isolating and selecting cells comprising a targeted gene modification. More particularly, the present invention discloses methods that enhance the recovery of cells carrying homologous recombination events. A main feature of the methods of the present invention is that expression of the selectable marker is regulated or under the control of the regulator. Upon homologous recombination, the regulator is not incorporated into the genome of the cell, allowing for expression of the selectable marker and selection of the desired cells.

The present invention represents a significant improvement over the currently available methods of generating cells comprising a disruption or modification of a target gene. Furthermore, the present invention provides an increase over previous technologies in both the speed and frequency at which homologous recombination events can be recovered.

The present invention also provides cells and animals that have been modified by the methods of the present invention to contain desired mutations or genomic modifications. In a preferred embodiment, the cells of the present invention are embryonic stem cells. In another preferred embodiment of the present invention, the animals are mice.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The terms "homologous" as used herein denotes a characteristic of a DNA sequence having at least about 70 percent sequence identity as compared to a reference sequence, typically

6

at least about 85 percent sequence identity, and preferably at least about 95 percent sequence identity as compared to a reference sequence. Most preferably, the homologous portions of the targeting vector will be 100% identical to the target DNA sequence. The percentage of sequence identity is calculated excluding small deletions or additions which total less than 25 percent of

5   the reference sequence. The reference sequence may be a subset of a larger sequence, such as a portion of a gene or flanking sequence, or a repetitive portion of a chromosome. However, the reference sequence is at least 18 nucleotides long, typically at least about 30 nucleotides long, and preferably at least about 50 to 100 nucleotides long.

"Disruption" or "modification" of a target gene or target sequence occurs when a

10   fragment of a DNA sequence locates and recombines with an endogenous homologous sequence. These sequence disruptions or modifications may include insertions, missense, frameshift, deletion, or substitutions, or replacements of DNA sequence, or any combination thereof. Insertions include the insertion of entire genes which may be of animal, plant, prokaryotic, or viral or other origin. Disruption or modification, for example, can alter or replace a promoter,

15   enhancer, or splice site of a target gene, and can alter the normal gene product by inhibiting its production partially or completely or by enhancing the normal gene product's activity.

The term, "transgenic cell", refers to a cell containing within its genome a specific gene that has been disrupted, modified, altered, or replaced completely or partially by the method of gene targeting.

20   As used herein, a "transgenic animal" is an animal that contains within its genome a specific gene that has been disrupted, modified, altered, or replaced completely or partially by the method of gene targeting. A transgenic animal includes both the heterozygote animal (*i.e.*, one defective allele and one wild-type allele) and the homozygous animal (*i.e.*, two defective alleles).

25   A "fragment" of a polynucleotide is a polynucleotide comprised of at least 9 contiguous nucleotides, preferably at least 15 contiguous nucleotides and more preferably at least 45 nucleotides, of coding or non-coding sequences.

A "host cell" includes an individual cell or cell culture which can be or has been a recipient for vector(s) or for incorporation of nucleic acid molecules and/or proteins. Host cells

30   include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent due to natural,

accidental, or deliberate mutation. A host cell includes cells transfected with the constructs and vectors of the present invention.

The term "homologous recombination" refers to the exchange of DNA fragments between two DNA molecules or chromatids at the site of homologous nucleotide sequences, *i.e.*, those sequences preferably having at least about 70 percent sequence identity, typically at least about 85 percent identity, and preferably at least about 90 percent identity, and most preferably 100 percent identity. Homology can be determined using a "BLASTN" algorithm, for example. It is understood that homologous sequences can accommodate insertions, deletions and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align.

As used herein, the term "target sequence" (alternatively referred to as "target gene sequence" or "target DNA sequence" or "target gene") refers to any nucleic acid molecule or polynucleotide of any gene to be modified by homologous recombination. The target sequence includes an intact gene, an exon or intron, a regulatory sequence or any region between genes.

As used herein, the term "regulator", refers to a sequence or sequences (*i.e.*, polynucleotide sequence or protein sequence) that regulates or controls expression of the selectable marker. The term "regulator" as used herein, excludes regulation of the expression of the selectable marker solely by degradation of RNA.

"Non-homologous integration" or "random integration", refers to the integration of DNA randomly and at any non-targeted genomic location. Non-homologous integration or random integration does not involve homologous recombination.

As used herein, the term "operably linked" includes reference to a functional linkage between a promoter and a nucleic acid sequence. The promoter sequence initiates and mediates transcription of the nucleic acid sequence.

As used herein, the term "promoter", generally refers to a regulatory region of DNA capable of initiating, directing and mediating the transcription of a nucleic acid sequence. Promoters may additionally comprise recognition sequences, such as upstream or downstream promoter or enhancer elements, which may influence the transcription rate.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a standard protocol for generating a transgenic animal. First, a targeting vector containing a selectable marker is created. Secondly, ES cells are transfected or

8

electroporated with the targeting vector and a drug such as G418 is added to select for the transfected or electroporated cells. Next, the cells are further analyzed for homologous recombination events. The transgenic animal is generated from a cell line in which homologous recombination has occurred.

5   Figure 2A and Figure 2B schematically depict and compare the DNA arrangements involved in homologous recombination and random integration of a targeting vector.

Figure 3A and Figure 3B illustrates the mechanism of positive-negative selection.

Figure 4A through Figure 4C depict and compare the various selection methods for identifying homologous recombination events in ES cells. Figure 4A depicts a traditional

10   positive selection method. Figure 4B depicts the positive-negative selection method. Figure 4C depicts the regulated positive selection method of the present invention.

Figure 5A and Figure 5B depict the general mechanism of the regulated positive selection method of the present invention.

Figure 6A through Figure 6D schematically depict the gene targeting vectors based on

15   the *lac* repressor system and display the changes in the DNA sequences that were introduced to generate these vectors. Figure 6A depicts the sequence for construct 3406 (c3406)(SEQ ID NO:13). Figure 6B depicts an example of a target gene with domains A-E. Figure 6C depicts the first-generation vector (Targeting Vector: PGK-*neo*) using a PGK-*neo* gene as a positive selection marker (SEQ ID NO:1). Figure 6D depicts the second-generation targeting vector.

20   The PGK-*lacO-neo* targeting vector contains the indicated base changes that introduce two *lacO* sites as well as a *Hind III* restriction enzyme site, as shown. The positions of the transcription start points (asterisks) and the methionine initiator codon (Met$_i$) is also noted. Partial sequence of the PGK promoter is shown (SEQ ID NO:2), with the bases that were deleted in the second-generation targeting vector (PGK-*lacO-neo*-NLS-*lacI*) marked with strikethrough font as shown

25   in Figure 6C. Figure 6E shows the final sequence of the DNA bases that encode the SV40-T antigen NLS from the methionine initiator codon of the NLS to the same codon of the *lac* repressor (SEQ ID NO:3).

Figure 7A and Figure 7B illustrate the mechanism of the present invention by which cells are selected for homologous recombination using the *lac* repressor system.

Figure 8 shows the sequences of oligonucleotides: 10164 (SEQ ID NO:4); 10165 (SEQ ID NO:5); 10218 (SEQ ID NO:6); 9959 (SEQ ID NO:7); 10219 (SEQ ID NO:8); and 4201 (SEQ ID NO:9), used to generate various constructs or vectors described in the foregoing examples.

Figure 9 schematically depicts four constructs used to test *lacI* repression of PGK-*lacO*-*neo* expression in mouse ES cells.

Figure 10 shows data relating to repression of PGK-*lacO-neo* expression in mouse ES cells. The colony number is graphed for each of the duplicate constructs that were tested at three different concentrations of G418.

Figure 11 schematically depicts three types of targeting vectors. The vectors all contain two gene-specific regions separated by the selectable marker, PGK-*lacO-neo*. "None" indicates the absence of a flanking gene; "*lacI* forw" and "*lacI* rev" indicate the presence of the *lacI* repressor expression cassette in the forward or reverse orientation. Both orientations express the *lac* repressor.

Figure 12 shows data relating to the recovery rate of homologous recombinants graphed for each target and each targeting construct. The numbers on top of the bar graphs represent the total numbers of colonies that were screened for homologous recombination.

Figure 13 shows data relating to NRSE regulation on the expression of a positive selection marker.

Figure 14A and Figure 14B show sequences for the Pst 1 and Pac 1 sites, as described in Example 1.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel compositions and methods useful in the production of cells and animals having within the genome a specific modification of a targeted gene. More particularly, the present invention is directed to various tools and methods that provide a fast, efficient, and reliable means of generating cells and animals comprising a specific genetic modification.

Construction of the targeting vector

The targeting vector or construct of the present invention may be produced using standard methods known in the art. (*See, e.g.*, Sambrook, *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; E.N. Glover (eds.), 1985, *DNA Cloning: A Practical Approach*, Volumes I and II;

M.J. Gait (ed.), 1984, *Oligonucleotide Synthesis*; B.D. Hames & S.J. Higgins (eds.), 1985, Nucleic Acid Hybridization; B.D. Hames & S.J. Higgins (eds.), 1984, *Transcription and Translation*; R.I. Freshney (ed.), 1986, Animal Cell Culture; Immobilized Cells and Enzymes, IRL Press, 1986; B. Perbal, 1984, A Practical Guide To Molecular Cloning; F.M. Ausubel *et al.*,

5   1994, *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.). For example, the targeting vector may be prepared in accordance with conventional ways, where sequences may be synthesized, isolated from natural sources, manipulated, cloned, ligated, subjected to *in vitro* mutagenesis, primer repair, or the like. At various stages, the joined sequences may be cloned, and analyzed by restriction analysis, sequencing, or the like.

10   The targeting vector or construct of the present invention typically comprises a first sequence homologous to a portion or region of a target gene sequence and a second sequence homologous to a second portion or region of the target DNA sequence. The targeting vector further comprises a selectable marker cassette comprising a sequence encoding a selectable marker, which is preferably positioned in between the first and the second DNA sequence that

15   are homologous to a region of the target DNA sequence. The targeting vector also comprises a sequence encoding a regulator, preferably, positioned outside of the first or second DNA sequence homologous to a region or portion of the target gene.

The targeting DNA can be constructed using techniques well known in the art. For example, the targeting DNA may be produced by chemical synthesis of oligonucleotides, nick-

20   translation of a double-stranded DNA template, polymerase chain-reaction amplification of a sequence (or ligase chain reaction amplification), purification of prokaryotic or target cloning vectors harboring a sequence of interest (*e.g.*, a cloned cDNA or genomic DNA, synthetic DNA or from any of the aforementioned combination) such as plasmids, phagemids, YACs, cosmids, bacteriophage DNA, other viral DNA or replication intermediates, or purified restriction

25   fragments thereof, as well as other sources of single and double-stranded polynucleotides having a desired nucleotide sequence. Moreover, the length of homology may be selected using known methods in the art. For example, selection may be based on the sequence composition and complexity of the predetermined endogenous target DNA sequence(s).

Preferably, the first and second sequences are of a functional component of a genomic

30   sequence to be targeted. Two fragments encoding separate portions of the target gene are generated. Although the size of each flanking region is not critical and can range from as few as

11

100 base pairs to as many as 100 kb, preferably each flanking fragment is greater than about 1 kb in length, more preferably between about 1 and about 10 kb, and even more preferably between about 1 and about 5 kb. Although larger fragments may increase the number of homologous recombination events in ES cells, larger fragments will also be more difficult to clone.

5   Typically, the portion of the gene included in the targeting construct is interrupted by insertion of a marker sequence (usually a selectable marker) that disrupts the reading frame of the interrupted gene so as to preclude expression of an active gene product. This most often causes a disruption (e.g., partial or complete inactivation) of normal production, structure, or function of the polypeptide encoded by the targeted gene of a single cell, selected cells or all of

10   the cells of an animal (or in culture).

When the targeting vectors of the present invention are introduced into embryonic stem cells, the transfected DNA can recombine with the target gene in the cell via the homologous sequences in both the vector and in the genomic region to be disrupted. The result of the homologous recombination event is often the insertion or incorporation of the selectable marker

15   sequence into an exon or portion of an exon of the target gene. Similarly, targeting constructs designed for knocking in genes can recombine at the homologous genomic site by homologous recombination and will result in the introduction of all or a portion of a gene into that locus. Techniques for knocking in genes are described in the art. (*See, e.g.*, Hanks *et al.*, 1995, *Science*, 269:679.

20   The selectable marker is a gene encoding a product that enables only the cells that carry the gene to survive and/or grow under certain conditions. A variety of selectable markers may be used in the practice of the present invention, including, for example, genes conferring resistance to compounds such as antibiotics, and genes conferring the ability to grow on selected substrates. In one aspect, the selectable marker is an antibiotic resistance gene such as the

25   neomycin resistance gene (neo) and the hygromycin resistance gene (hyg). (*See, e.g.*, Southern, P., and P. Berg, 1982, *J. Mol. Appl. Genet.* 1:327-341; Te Riele, H., *et al.*, 1990, *Nature* 348:649-651). Selectable markers that may be used in accordance with the present invention are described in the art. (*See, e.g.*, Sambrook, J., et al., 1989, *Molecular Cloning--A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., Chapter 16). In many cases

30   it is desirable to disrupt genes by positioning the positive selection marker in an exon, *i.e.*, a functional component, of a gene to be disrupted or modified.

12

The regulator inhibits or suppresses expression of the selectable marker, and is removed upon homologous recombination, but retained upon random integration of the targeting vector. Various genetic elements are incorporated into the regulator allowing it to control expression of the selectable marker. In one aspect, the regulator comprises sequences that regulate or control the expression of the selectable marker at any step in the gene expression pathway, for example, at the point of transcription. In accordance with this aspect, the targeting vector may be comprised of a transcription control system, such as an operator/repressor system, for instance. In this construction of the targeting vector, the regulator comprises sequences that interact with or bind to sequences present within the selectable marker cassette preventing or repressing expression of the selectable marker. Other suitable transcriptional control systems capable of regulating expression of the selectable marker may be used in accordance with the present invention.

The regulator may also be comprised of elements that control expression of the selectable marker at the steps of transcription, pre-mRNA processing (i.e., splicing, polyadenylation, capping), mRNA transport, mRNA stability, translation, protein stability, and protein activity. The regulator may also comprise other sequences or DNA binding proteins that affect degradation or localization of the selectable marker or sequences, for example, a nuclear localization signal. (*See, e.g.*, Hannan *et al.*, *Gene* 130(2):233-239). The regulator may also comprise sequences that direct or enhance its expression including, promoters, polyadnelyation signals, introns, and the like.

In one aspect, the selectable marker cassette comprises a selectable marker gene linked to a sequence that activates transcription of the selectable marker. In this aspect, preferably, the selectable marker cassette comprises a promoter sequence operably linked to the sequence encoding the selectable marker. The selectable marker cassette may also comprise other regulatory sequences. For example, the promoter sequence may further comprise at least one operator sequence placed adjacent to or within the promoter sequence. In this construction, the regulator interacts with or binds to the promoter/operator sequence to regulate expression of the selectable marker. In accordance with this aspect, the regulator comprises a repressor sequence compatible with the promoter/operator sequence to inhibit or repress expression of the selectable marker of the targeting vector.

13

A preferred design of the targeting vector includes a selectable marker cassette positioned in between the first and second sequence homologous to a portion or region of the target gene. The selectable marker cassette comprises a promoter region operably linked to a sequence encoding the selectable marker. Preferably, the selectable marker cassette further comprises at least one operator site placed adjacent to or within the promoter. In a preferred embodiment, the promoter region comprises a PGK promoter sequence and at least one operator site, and the selectable marker is the *neo* gene. The regulator is preferably positioned outside of and adjacent to the first or second sequence homologous to the target gene and interacts with or binds to sequences (i.e., regulatory binding sites) within the promoter region to repress or inhibit expression of the selectable marker.

In one aspect, the selectable marker is controlled by a *lac* operator/repressor system. In this design, the targeting vector comprises a selectable marker cassette comprising a promoter sequence, at least one *lac* operator sequence, and a sequence encoding a selectable marker, preferably, positioned in between the first and second sequences homologous to a region or portion of the target DNA. In a preferred aspect, the promoter region comprises the PGK promoter and two *lac* operator sequences positioned next to or within the PGK promoter sequence. The regulator is preferably positioned outside either the first or second sequences homologous to the target gene, and comprises a *lac* repressor sequence. In a preferred embodiment, the regulator also comprises sequences corresponding to a nuclear localization signal (NLS), resulting in a regulator that comprises sequences encoding a *lac* repressor and a nuclear localization signal. In a preferred embodiment, the NLS originates from the simian virus 40 large-T antigen. (*See, e.g.,* Hu and Davidson, 1991, *Gene* 99:141-150). An example of this targeting vector is shown in Figure 6.

Any promoter system available in the art may be used in the practice of the present invention. Examples of such promoters include the beta.-lactamase (penicillinase) system, a tryptophan (trp) promoter system, and the like. (*See, e.g.,* Chang *et al.,* 1978, *Nature,* 275: 615; Itakura, *et al.,* 1977, *Science,* 198: 1056; Goeddel *et al.,* 1979, *Nature* 281: 544; Goeddel, *et al.,* 1980, *Nucleic Acids Res.* 8: 4057; Siebenlist, *et al.,* 1980, *Cell* 20: 269).

Any element capable of regulating the expression of the selectable marker may be used in accordance with the present invention. Thus, the regulator may be comprised of elements other than a DNA sequence encoding a protein. The present invention contemplates that expression of

14

the selectable marker is regulated at any step in the gene expression pathway. For example, the regulator could act in *cis*, for example, as a transcriptional silencer element such as NRSF/Rest, REST, MeCP2, NRF, rGH, NRE and COL4. (*See, e.g.*, Chen *et al.*, 1998, *Nat. Gen*; Chang *et al.*, 1995, *Cell* 80:949-957; Xinshen-Nan *et al.*, 1997, *Cell* 88:471-481; Nourkakhsh *et al.*, 1997,

5   *Immunbiolo.*; Roy *et al.*, 1994, *Eur. J. Biochem.*; Li-Weber *et al.*, 1993, *J. of Immunology*; Hanel *et al.*, 1995, *JBC*).

Other DNA sequences or proteins that affect the uptake of the targeting vector after introduction into the cells may also be present. For example, sequences or DNA binding proteins that affect degradation or localization of the vector following entry into the targeted

10   cells or molecules that affect the catalysis of homologous recombination may be incorporated in the targeting vector of the present invention. Moreover, other regulatory sequences may be incorporated into the targeting vector to disrupt or control expression of a particular gene in a specific cell type.

In a preferred embodiment, the targeting vector(s) of the present invention is generated in

15   two steps. The first step involves generating a first vector comprised of a first sequence homologous to a region or portion of the target gene sequence, a second sequence homologous to a region or portion of the target gene sequence, and a third sequence that encodes a selectable marker. In the second step, standard subcloning methods known in the art may be used to incorporate the regulator into the targeting vector.

20   In another aspect of the present invention, a plasmid comprising: a first gene-specific region of homology; the insert containing the selectable marker, for example, a PGK-*lac* operator-selectable cassette; and a second gene-specific region of homology is generated. Standard subcloning methods are used to insert the regulator gene, such as a NLS-*lac*I sequence, into the vector. In a preferred embodiment, the selectable marker and the regulator are separated

25   by at least one region of homology. For example, the regulator may be placed outside of and adjacent to the first or second sequence substantially homologous to the target gene.

In a preferred embodiment, the method comprises producing a targeting vector comprising a *lac* repressor system. As depicted in Figure 6, a first-generation vector is produced using a PGK-neo gene as a positive selection marker. A second-generation targeting vector is

30   produced and comprises a partial sequence of the PGK promoter containing the indicated base changes that result from introducing two *lacO* sites, in addition to a *Hind III* restriction enzyme

15

site. The positions of the transcription start points (asterisks) and the methionine initiator codon (Met$_i$) is noted in Figure 6. A regulator comprising a sequence encoding the SV40-T antigen NLS from the methionine initiator codon of the NLS and a *lac* repressor is subcloned into this PGK-*lacO-neo* targeting vector as indicated in Figure 6. The resulting targeting vector

5 comprises a first and second sequence homologous to the target gene, a positive selection marker comprising a PGK-*lacO-neo* sequence, and a regulator comprising a NLS and *lac* repressor sequence.

In another embodiment of the present invention, the targeting vector is prepared directly from a plasmid genomic library using the methods described in pending U.S. Patent Application

10 No.: 08/971,310, filed November 17, 1997, the disclosure of which is incorporated herein in its entirety. Generally, a sequence of interest is identified and isolated from a plasmid library in a single step using, for example, long-range PCR. Following isolation of this sequence, a second polynucleotide that will disrupt the target sequence can be readily inserted between two regions encoding the sequence of interest. The regulator is subsequently subcloned into the vector.

15 In accordance with this embodiment, the targeting vector or construct is generated using ligation-independent cloning to insert two different fragments of the homologous sequence into a vector having a selectable marker cassette comprising the selectable marker gene positioned between the two different homologous sequence fragments in the construct. In one aspect of this embodiment, the homologous sequences may be obtained by: generating two primers

20 complementary to the target; annealing the primers to complementary sequences in a mouse genomic DNA library containing the target region; and amplifying sequences homologous to the target region. The products of the amplification reaction, which have endpoints formed by the primers, are then isolated. Preferably, amplification is by PCR; more preferably, amplification is by long-range PCR.

25 Applying this method of generating the targeting vector, the present invention obviates the need for hybridization isolation, restriction mapping, and multiple cloning steps. For example, a short sequence can be used to design oligonucleotide probes that can be directly amplified to create the targeting vector. For example, a short sequence (*e.g.*, EST) can be used to design oligonucleotide probes. These probes can be used in the direct amplification procedure to

30 create constructs or can be used to screen genomic or cDNA libraries for longer full-length

genes. Thus, it is contemplated that any gene can be quickly and efficiently prepared using the methods of the present invention for use in producing cells having a targeted gene modification.

Production and Selection of Cells Comprising a Targeted Gene Modification

Once an appropriate targeting vector(s) has been prepared, the vector may be introduced into an appropriate host cell using any method known in the art. Various techniques may be employed in the present invention, including, for example, pronuclear microinjection; retrovirus mediated gene transfer into germ lines; gene targeting in embryonic stem cells; electroporation of embryos; sperm-mediated gene transfer; and calcium phosphate/DNA co-precipitates, microinjection of DNA into the nucleus, bacterial protoplast fusion with intact cells, transfection, polycations, *e.g.*, polybrene, polyornithine, *etc.*, or the like (*See, e.g.*, U.S. Pat. No. 4,873,191; Van der Putten, *et al.*, 1985, *Proc. Natl. Acad. Sci., USA* 82:6148-6152; Thompson, *et al.*, 1989, *Cell* 56:313-321; Lo, 1983, *Mol Cell. Biol.* 3:1803-1814; Lavitrano, *et al.*,1989, *Cell*, 57:717-723). Various techniques for transforming mammalian cells are known in the art. (*See, e.g.*, Gordon, 1989, *Intl. Rev. Cytol.*, 115:171-229; Keown *et al.*, 1989, *Methods in Enzymology*; Keown *et al.*, 1990, *Methods and Enzymology*, Vol. 185, pp. 527-537; Mansour *et al.*, 1988, *Nature*, 336:348-352).

In one aspect, the targeting vector is introduced into host cells by electroporation. In this process, electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the vector. The pores created during electroporation permit the uptake of macromolecules such as DNA. (*See, e.g.*, Potter, H., *et al.*, 1984, *Proc. Nat'l. Acad. Sci. U.S.A.* 81:7161-7165).

Any cell type capable of homologous recombination may be used in the practice of the present invention. Examples of such target cells include cells derived from vertebrates including mammals such as humans, bovine species, ovine species, murine species, simian species, and ether eucaryotic organisms such as filamentous fungi, and higher multicellular organisms such as plants.

Preferred cell types are embryonic stem (ES) cells, which are typically obtained from pre-implantation embryos cultured *in vitro*. (*See, e.g.*, Evans, M. J., *et al.*, 1981, *Nature* 292:154-156; Bradley, M. O., *et al.*, 1984, *Nature* 309:255-258; Gossler *et al.*, 1986, *Proc. Natl. Acad. Sci. USA* 83:9065-9069; and Robertson, *et al.*, 1986, *Nature* 322:445-448). The ES cells are cultured and prepared for introduction of the targeting vector using methods well known to the skilled

artisan. (*See, e.g.,* Robertson, E. J. ed. "Teratocarcinomas and Embryonic Stem Cells, a Practical Approach", IRL Press, Washington D.C., 1987; Bradley *et al.*, 1986, *Current Topics in Devel. Biol.* 20:357-371; by Hogan *et al.* in "Manipulating the Mouse Embryo": A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y., 1986; Thomas *et al.*, 1987, *Cell*

5   51:503; Koller *et al.*, 1991, *Proc. Natl. Acad. Sci. USA*, 88:10730; Dorin *et al.*, 1992, *Transgenic Res.* 1:101; and Veis *et al.*, 1993, *Cell* 75:229). The ES cells that will be inserted with the targeting vector are derived from an embryo or blastocyst of the same species as the developing embryo into which they are to be introduced. ES cells are typically selected for their ability to integrate into the inner cell mass and contribute to the germ line of an individual when

10   introduced into the mammal in an embryo at the blastocyst stage of development. Thus, any ES cell line having this capability is suitable for use in the practice of the present invention.

The present invention may also be used to knockout genes in other cell types, such as stem cells. By way of example, stem cells may be myeloid, lymphoid, or neural progenitor and precursor cells. These cells comprising a disruption or knockout of a gene may be particularly

15   useful in the study of target gene function in individual developmental pathways. Stem cells may be derived from any vertebrate species, such as mouse, rat, dog, cat, pig, rabbit, human, non-human primates and the like.

After the targeting vector has been introduced into cells, the cells where successful gene targeting has occurred are selected. Insertion of the targeting vector into the targeted gene is

20   typically detected by selecting cells for expression of the marker gene. The cells transformed with the targeting vector of the present invention are subjected to treatment with an appropriate agent that selects against cells not expressing the selectable marker. Only those cells expressing the selectable marker gene survive and/or grow under certain conditions. For example, cells that express the introduced neomycin resistance gene are resistant to the compound G418, while cells

25   that do not express the neo gene marker are killed by G418. The targeting vector of the present invention is constructed so that the regulator is disposed of or degraded by the cell upon homologous recombination, and thus, expression of the selectable marker is permitted. Upon random integration, substantially all of the targeting vector, including the regulator, may be incorporated into a random site in the genome of the cell and expression of the selectable marker

30   is inhibited or repressed by the regulator.

Integration of the transfected DNA into the appropriate site of the genome results in the stable acquisition and expression of the selectable marker, wherein the first and second DNA sequences of the targeting vector are incorporated within the homologous portions of the endogenous target DNA of the cell. The targeting vector is constructed, so that upon

5   homologous recombination, the regulator is not incorporated into the genome of the cell. Non-incorporation of the regulator allows expression of the selectable marker, and thus, identification of cells, wherein gene targeting has occurred. Predominantly, however, integration of the transfected DNA occurs at a random site in the genome of the cell. When random integration occurs, the targeting vector including the regulator is inserted into a random site in the genome

10   of the cell. As expression of the selectable marker is under the control of the regulator, the cells wherein random integration occurs do not survive addition of the selective agent, as the regulator incorporated into the cell-blocks or inhibits expression of the marker gene.

As illustrated in Figure 7, upon homologous recombination, *lac* repressor inhibition of *neo* transcription is relieved upon homologous recombination. The cells expressing the

15   selectable marker can be identified through the addition of a drug, such as G418. Conversely, upon random integration, the regulator is incorporated into a random site in the genome of cells and thus, retains the ability to inhibit or suppress expression of the selectable marker. As a result of random integration of the targeting vector, the regulator interacts with the promoter operably linked to the selectable marker to inhibit transcription of the selectable marker gene. Addition of

20   the selection agent kills these cells. More specifically, after using electroporation to place the vectors into cultured ES cells, neomycin was added to the culture medium to select for the growth of cells expressing the *neo* gene. Expression of the *neo* gene requires that: (1) the cell was successfully electroporated; and (2) *lac* repressor inhibition of *neo* transcription was relieved, *i.e.,* by homologous recombination. This vector is then introduced into ES cells where

25   a single positive selection selects for transfected cells and enriches the population for clones derived from the desired homologous recombination event as described below.

Successful recombination may be identified by analyzing the DNA of the selected cells to confirm homologous recombination. Various techniques known in the art, such as PCR and/or Southern analysis may be used to confirm homolgous recombination events.

30   The PCR screening procedure uses a target gene specific oligonucleotide that is not present on the targeting vector and an oligonucleotide corresponding to sequences in the

19

selectable marker cassette. Oligonucleotides outside the targeting vector are used to differentiate homologous recombinants from random integrations of the targeting vector. In general, oligonucleotides not present on the targeting vector are tested on wild type ES cell DNA in combination with target gene-specific oligonucleotides that are adjacent to the insertion site of

5   the selectable marker cassette. Oligonucleotides producing background bands or failing to give the predicted size product are eliminated. A single target gene-specific oligonucleotide is selected and paired with an oligonucleotide corresponding to sequences in the selectable marker cassette. ES cells that are PCR positive in this screen are confirmed by a second PCR experiment that utilizes a different pair of target gene-specific and selectable marker-specific

10   oligonucleotides that are adjacent to, but distinct from, the original oligonucleotide pair. In addition, this protocol may be repeated using oligonucleotides specific for target gene sequences located on the opposite side of the selectable marker in conjunction with a marker-specific oligonucleotide. In this way proper integration (i.e., homologous recombination) of both homologous sequences of the targeting vector is verified.

15   Southern analysis may also be used to confirm the ES cell targeting event. A unique probe that is external to the targeting sequences themselves is developed and used to screen by Southern analysis. The probe should not contain any repetitive DNA elements and can be upstream or downstream from the targeting construct. The probe can be used in conjunction with Southern analysis of each ES clone to determine whether or not a targeting event has occurred.

20   In addition to defining a homologous recombination DNA fragment, Southern analysis also allows for assessment of the ratio of mutant to wild-type bands, and thus an assessment of whether the ES line is a pure, clonally-derived population.

Production of Genetically Altered Animals

25   Embryonic stem cells which have been modified can be injected into the blastocoel of a blastocyst and grown in the uterus of a pseudopregnant female. In order to readily detect chimeric progeny, the blastocysts can be obtained from a different parental line than the embryonic stem cells. For example, the blastocysts and embryonic stem cells may be derived from parental lines with different hair color or other readily observable phenotype. The resulting

30   chimeric animals can be bred in order to obtain non-chimeric animals which have received the modified genes through germ-line transmission. Techniques for the introduction of embryonic

20

stem cells into blastocysts and the resulting generation of chimeric animals are well known. (*See e.g.*, Bradley, A. "Production and analysis of chimeric mice", pp. 113-151 in Robertson, E. (ed.), *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Oxford IRL Press (1987); and Hogan, B., *et al.*, 1986, Manipulating the Mouse Embryo, Cold Spring Harbor, N.Y.).

5   An alternate method of preparing an embryo containing ES cells that possess the targeting vector is to generate "aggregation chimeras". A morula of the proper developmental stage (about 2 1/2 days post-fertilization for mice) is isolated. The zona pellucida can be removed by treating the morula with a solution of mild acid for about 30 seconds, thereby exposing the "clump" of cells that comprise the morula. Certain types of ES cells such as the R1 cell line for mice can then be co-cultured with the morula cells, forming an aggregation chimera embryo of morula and ES cells. (*See, e.g.*, Joyner, A. L., 1993, *Gene Targeting, The Practical Approach Series*, JRL Press Oxford University Press, New York).

If animals homozygous for the targeted mutation are desired, they can be prepared by crossing animals heterozygous for the targeted mutation. Mammals homozygous for the disruption may be identified by Southern blotting of equivalent amounts of genomic DNA from mammals that are the product of this cross, as well as mammals of the same species that are known heterozygotes, and wild-type mammals. Alternatively, specific restriction fragment length polymorphisms can be detected which co-segregate with the mutant locus. Probes may be designed to screen the Southern blots for the presence of the targeting construct in the genomic DNA. In addition, PCRs can be used to genotype animals as wild-type, heterozygous mutant or homozygous mutant.

Other means of identifying and characterizing the offspring having a disrupted gene are also available. For example, Northern blots can be used to probe mRNA obtained from various tissues of the offspring for the presence or absence of transcripts. Differences in the length of the transcripts encoded by the targeted gene can also be detected. In addition, Western blots can be used to assess the level of expression of the targeted gene by probing the Western blot with an antibody against the protein encoded by the targeted gene. Protein for the Western blot may be isolated from tissues where this gene is normally expressed. Finally, *in situ* analysis (such as fixing the cells and labeling with antibody or nucleic acid probe) and/or FACS (fluorescence activated cell sorting) analysis of various cells from the offspring can be conducted using suitable antibodies to look for the presence or absence of the gene product.

21

<u>Advantages</u>

The present invention employs a regulated positive selection method that provides significant advantages over conventional methods of producing cells and animals comprising a targeted gene modification. The following compares two widely used methods of producing

5   knockout cells and knockout animals with the regulated positive selection method of the present invention.

As illustrated in Figure 3, the PNS method involves a two-step cell culturing process consisting of a positive selection step and a negative selection step. In the PNS process, a second drug, in addition to neomycin, is added that kills cells as a direct consequence of expression of

10   the negative selection marker (See Figure 4 and Table 1). Although this process adds to the recovery of homologous recombinants, the PNS method presents two important drawbacks. First, the two-step process may be time-consuming and laborious, and second, the addition of a second drug, such as FIAU and related drugs may hinder the ability of ES cells to populate mice and transmit the targeted allele through the germline. For example, most targeting vectors

15   employed in the PNS method use both PGK-Neo and HSV-TK to perform positive and negative selection, respectively. However, gancyclovir treatment of ES cells is known to be quite toxic, and may negatively affect the ability of ES cells to generate animals (i.e., chimeric mice) and/or to subsequently populate the germline of these animals. Moreover, cells comprising random integration events will also inactivate expression of the negative selection marker, allowing these

20   cells to remain present in the cell population.

Significant advantages are presented by the regulated positive method of the present invention for producing or identifying cells having a targeted gene modification as compared to the traditional positive selection method (Figure 1) and the PNS method. The method of the present invention represents a significant improvement over both the traditional positive selection

25   and PNS methods as the methods of the present invention enrich the cell population for homologous integration events while employing only a single drug in a one-step positive selection. Importantly, the method of the present invention allows for the selection of transfected cells and the enrichment for homologous recombinants to occur in one step with the addition of a single drug, *i.e.*, no negative selection applied. The advantages of the methods of the present

30   invention over the traditional positive selection method and the PNS method are summarized in the following Table I:

22

Selection Type	Number of Steps	Number of Drugs	Screening	Side-Effects	Time/ Efforts
Positive	1	1	Many Colonies	None known	Slow/labor intensive: many colonies must be screened
PNS	2	2	Fewer Colonies	Negative selection hinders chimera generation and germline transmission; toxicity	Slow: Requires 2 drugs and is often a 2 step process
Regulated Positive Selection	1	1	Fewer Colonies	None known	Faster/Easier: 1 step and fewer colonies to screen

TABLE I

5   The regulated positive selection method of the present invention clearly reveals increases over

previous technologies in both the speed and frequency at which homologous recombination events can

be recovered. Moreover, restricting expression of the positive selection marker to clones carrying the

homologous recombination event provides a powerful means to enhance the recovery of the desired

mutant cell lines without the need for additional drugs, selections, screens or cell manipulations beyond

10   those used in the standard positive selection. Thus, the present invention provides a method that is much

more rapid and efficient than currently-employed processes.

As described herein, one of the most restrictive bottlenecks in generating animals

comprising a targeted gene modification is the identification and isolation of the rare cell line

carrying the homologous recombination event. The present invention represents a significant

15   improvement over the currently available methods of producing modified cells and animals

having a disruption of a target gene by enriching the cell population for homologous integration

events. One of the significant advantages of the present invention is that it substantially reduces

the number of colonies that need be screened to identify cell lines containing a desired genetic

modification. Using conventional methods, a number of random integration events would still

20   survive and grow under positive selection. The methods of the present invention markedly

23

reduce the number of random integration events that would normally grow under positive selection, thus, providing a more rapid and efficient process in generating cells with targeted gene modifications.

More particularly, the present invention provides methods that enhance the recovery of cell lines carrying homologous recombination events by controlling the expression of the positively-selected marker gene. Specifically, genetic elements that down regulate expression of the marker gene are cloned into the plasmid DNA adjacent to the regions that share homology with the target sequences. Homologous recombination removes these elements, which in turn increases expression of the marker gene and enhances the identification of homologous recombination events. Thus, the present invention provides fast, efficient, and reliable methods of generating cells and animals comprising a targeted gene modification.

EXAMPLES

The following examples are provided solely to illustrate the claimed invention. The present invention, however, is not limited in scope by the exemplified embodiments, which are intended as illustrations of single aspects of the invention only, and methods functionally equivalent are within the scope of the invention. Various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

Example 1: Targeting Vector Construction

Generation of the PGK-lacO-neo Gene. The PGK-lacO-neo hybrid gene was generated in the following manner: Using pDG2 (see U.S. Patent Application No.: 08/971,310, filed November 17, 1997) as a template, oligonucleotides 10218 and 9959 (Figure 8) were used in the polymerase chain reaction (PCR) using Expand polymerase (Roche Biochemicals) to generate a DNA fragment containing the second *lacO* site (Figure 6). This fragment was digested with *Hind III* and *NcoI* (all restriction enzymes from New England Biolabs, Beverly, MA). In parallel, the same reaction conditions---except using oligonucleotides 10219 and 4201--- were used to generate another DNA fragment containing the first *lacO* site (Figure 6). This fragment was digested with *Hind III* and *Eco RI*. The two PCR fragments were then ligated together into the *Nco I* and *Eco RI* sites of pDG2, replacing the wild-type sequence between these restriction sites. This plasmid was designated as construct 3363.

24

Generation of NLS-lacI Gene. The NLS-*lacI* gene was generated in the following manner: Using plasmid pTrcHisA (Invitrogen, Carlsbad, CA) as a template, oligonucleotides 10164 and 10165 (Figure 8) were used in the polymerase chain reaction (PCR) using Expand polymerase (Roche Biochemicals) to generate a DNA fragment containing the *lacI* gene. The

5   cycling conditions followed the supplier's recommendations and were as follows: 25 cycles at 94°C for 10 seconds, 50°C for 30 seconds and 68°C for 70 seconds. These cycles were preceded by one denaturation heating at 94°C for 2 minutes and were followed by an incubation at 68°C for 7 minutes. The PCR fragment was digested with *Eco RI* and then subcloned into the *Eco RI* sites of pCX-EGFP (see Hadjantonakis et al., 1998, *Mech Dev* 76:79-90), generating construct

10   3359. Construct 3361 was also made, identical to c3359 except that the NLS-*lacI* gene is present in the reverse orientation. Finally, c3359 was digested with *Sal I* and *Hind III*, the DNA ends were made blunt using T4 DNA polymerase (Roche Biochemicals), and the DNA fragment containing NLS-*lacI* and the surrounding enhancer, promoter, intron and polyadenylation sequences was subcloned into the *Pst I* and *Pac I* sites (*See* Figure 14A and 14B) to generate

15   constructs c3406 (Figure 6A) and c3408, which are identical except that each contains the entire *lac* repressor expression cassette in opposite orientations.

    Targeting vector comprising lac repressor system. A targeting vector based on a *lac* repressor system is illustrated in Figure 6. Changes in the DNA sequences that were introduced to generate these vectors are shown. The first-generation vector (Targeting Vector: PGK-*neo*)

20   uses a PGK-*neo* gene as a positive selection marker. Partial sequence of the PGK promoter is shown, with the bases that were deleted in the second-generation targeting vector (PGK-*lacO-neo*-NLS-*lacI*) marked with strikethrough font. The PGK-*lacO-neo* targeting vector contains the indicated base changes that introduce two *lacO* sites as well as a *Hind III* restriction enzyme site. The positions of the transcription start points (asterisks) and the methionine initiator codon

25   (Met$_i$)is also noted. The final sequence lists the DNA bases that encode the SV40-T antigen NLS from the methionine initiator codon of the NLS to the same codon of the *lac* repressor.

Example 2: Repression of PGK-*lacO-neo* Expression in Mouse ES Cells

    Four test vectors identified as c3400, c3398, c3396, and c3394 were created to test repression of the selectable marker. Each vector was identical to the other except for the

30   presence or absence of the *lacO* or *lacI* sequences (Figure 9). The *lacI* sequences were ligated together with the selectable marker sequences using the *Bam HI* and *Sal I* restriction sites present

in each parent vector, using the plasmid backbone from the selectable marker plasmid. The wild-type PGK-*neo* fragment was derived from pDG-2; PGK-*lacO-neo* from c3363; *lacI* in the coding orientation from c3359; and *lacI* in the non-coding orientation from c3361.

To determine whether NLS-*lacI* could repress expression of PGK-*lacO-neo* and thus decrease the number of random integration events recovered, the four constructs outlined in Figure 9 were introduced into ES cells. The effects on *neo* expression were assessed by counting G418-resistant colonies. Importantly, these constructs are identical, except for the presence of *lacO* sites and whether NLS-*lacI* is cloned in the coding or non-coding orientation. By limiting the changes in the plasmids to those sequences involved in *lac* repression, any observed effects in *neo* expression can be directly attributed to the specifically introduced *lacO* or *lacI* sequences as opposed to general changes in the vector backbone or other differences outside of the *lac*-related sequences.

The basic protocol was as follows: the constructs were digested with *Swa I* to generate linear DNA. As a control for experimental variability, duplicate constructs (for each of those listed in Figure 9) were prepared and tested in parallel on separate days. The digested plasmids were resuspended in distilled water to a concentration of 1 µg/µl and introduced into mouse ES cells using electroporation. Rapidly growing ES cells were trypsinized to make single cell suspensions. The respective targeting vectors were linearized with a restriction endonuclease and 2 µg of DNA was added to 10 x 10⁶ ES cells in ES medium {High Glucose DMEM (without L-Glutamine or Sodium Pyruvate) with LIF (Leukemia Inhibitory Factor-Gibco 13275-029 "ESGRO") 1,000 units/ml, and 12% Fetal Calf Serum}. Cells were placed into a 2 mm gap cuvette and electroporated on a BTX electroporator at 400 µF resistance and 200 volts. Subsequently, the cells were plated using G418 concentrations of 150 µg/ml, 200 µg/ml or 400 µg/ml. After 10-12 days of selection, the total number of G418-resistant colonies were counted.

The *lacI* or *lacO* sequences alone (c3398, c3396 compared to c3400) resulted in a decrease in colony number at each concentration of G418 (Figure 10). The *lacO* and *lacI* sequences together (c3394) also reduced the number of G418-resistant colonies. However, this reduction differed from those that resulted from *lacO* or *lacI* alone in two important ways. First, the reduction observed with c3394 was significantly larger than were the reductions observed from c3398 and c3396, particularly at the higher G418 concentrations. This result suggests that the *lacO* and *lacI* sequences act in concert to down regulate *neo* expression, as would be

26

expected for a regulatory system dependent on formation of the *lac* operator-repressor complex. Second, the c3394-dependent reduction was enhanced at higher concentrations of G418, whereas the other reductions were not. This observation indicates that the *lac* repressor effectively down regulates PGK-*lacO-neo* expression, but does not completely block it. Thus, cells transfected

5   with PGK-*lacO-neo* and expressing the *lac* repressor appear to express *neo* at a low level; at low concentrations of G418, this level of *neo* expression appears to be enough to support growth whereas at higher concentrations of G418 it is not. Taken together, the results for this experiment indicate that the *lac* repressor can inhibit PGK-*lacO-neo* expression in mouse ES cells, and in so doing, reduce the number of random integration events that grow under positive

10   selection.

Example 3: Enhancement of Recovery of Homologous Recombination Events

To determine whether the *lac* repressor system could be used to enhance the rate of recovery of homologous recombination events, three different types of targeting vectors were constructed (Figure 11) and used to direct homologous recombination to six different target

15   genes. These genes belonged to different gene families: serine protease, metalloprotease, serine/threonine kinase, serine protease inhibitor, G-protein-coupled receptor, and acylphosphatase.

The results outlined in Figure 12 clearly demonstrate that a repressor system can be used to enhance the rate at which homologous recombinants are recovered. Comparing the rates that

20   were observed using no flanking gene ("none") to those obtained using *lacI* forward or reverse ("*lacI* forw + rev") reveals a higher rate for "*lacI* forw + rev" in five of the targets. The enhancement varied from approximately two- to six-fold, and in one case (T667), no homologous recombinants were detected unless the *lac* system was employed. In the sixth case (T752), the rates using "none" and "*lacI* forw + rev" were essentially equal. This target also

25   displayed the highest recovery rate relative to the other five targets, suggesting that it may represent a recombination "hotspot" where a rate enhancement was not needed to easily detect a homologous recombination event. In summary, the results from this example reveal that the *lac* operator-repressor system can significantly improve upon existing methods for making targeted gene disruptions in mouse ES cells.

30   **Example 4: Regulation of the Selectable Marker with a Silencer Element**

27

Three copies of the NRSE silencer element derived from the S36 region of the SCG10 gene (Schoennherr and Anderson, Science, 1995) were subcloned into the *Hind-III* site of c319. This *Hind-III* site is positioned near the PGK promoter region. The sequence of the silencer element is: cagaggcactctccgtggtgctgaaa (SEQ ID NO:10)

5   The oligos used for cloning into the *Hind-III* site are the following SEQ ID NO:11 and SEQ ID NO:12. The silencer regions for both sequences are highlighted.

AGCTT**tttcagcaccacggagagtgcctctg**CT**tttcagcaccacggagagtgcctctg**CT**tttcagcaccacggagag**tgcc tctgA (SEQ ID NO:11)

AGCTT**cagaggcactctccgtggtgctgaaa**AG**cagaggcactctccgtggtgctgaaa**AG**cagaggcactctccgtggtg**

10 **ctgaaa**A (SEQ ID NO:12)

The number of ES cell clones which survived G418 selection between the control construct (c319) and the construct with the 3 copies of the silencer element (c 2650). Three different DNA concentrations were used under standard electroporation conditions. The concentrations were 5, 15, and 30 ug DNA.

15 The number of colonies after G418 selection is shown in the following Table 2:

Table 2:

Construct	5 ug	15 ug	30 ug
PGK-Neo	1536	1064	2180
NRSE-PGK-NEO	336	604	1848

As shown in Table 2 and Figure 13, there was a 78% decrease in colonies from the NRSE

20 construct compared to the control construct at 5ug DNA concentration; a 43% decrease in colonies from the NRSE construct compared to the control construct at 15ug DNA concentration; and a 15% decrease in colonies from the NRSE construct compared to the control construct at 30ug DNA concentration.

The relative increase in colony number with increasing DNA concentration may be the

25 result of an increase in copy number or higher frequency of tandem integration events which would lead to higher levels of expression from the PGK promoter.

It is understood that the present invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents, etc., described herein, as these may vary. It is also to be understood that the terminology used herein is used for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention. Preferred

5   methods, devices, and materials are described, although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention. All references cited herein are incorporated by reference herein in their entirety.